Testing a Spectral Model of Tonal Affinity with Microtonal Melodies and Inharmonic Spectra

Testing a Spectral Model of Tonal Affinity with Microtonal Melodies and Inharmonic Spectra

Andrew J. Milne
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, 2751, NSW, Australia

Robin Laney
The Open University, Department of Computing and Communications, Milton Keynes, MK7 6AA, UK

David B. Sharp
The Open University, Department of Engineering and Innovation, Milton Keynes, MK7 6AA, UK

Author Note

Dr. Andrew J. Milne, MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, 2751, NSW, Australia.
a.milne@uws.edu.au

Abstract

Tonal *affinity* is the perceived goodness of fit of successive tones. It is important because a preference for certain intervals over others would likely influence preferences for, and prevalences of, "higher-order" musical structures such as scales and chord progressions. We hypothesize that two psychoacoustic (spectral) factors—harmonicity and spectral pitch similarity—have an impact on affinity. The *harmonicity* of a single tone is the extent to which its partials (frequency components) correspond to those of a harmonic complex tone (whose partials are a multiple of a single fundamental frequency). The *spectral pitch similarity* of two tones is the extent to which they have partials with corresponding, or close, frequencies. To ascertain the unique effect sizes of harmonicity and spectral pitch similarity, we constructed a computational model to numerically quantify them. The model was tested against data obtained from 44 participants who ranked the overall affinity of tones in melodies played in a variety of tunings (some microtonal) with a variety of spectra (some inharmonic). The data indicate the two factors have similar, but independent, effect sizes: in combination, they explain a sizeable portion of the variance in the data (the model-data squared correlation is $r^2 = .64$). Neither harmonicity nor spectral pitch similarity require prior knowledge of musical structure, so they provide a potentially universal bottom-up explanation for tonal affinity. We show how the model—as optimized to these data—can explain scale structures commonly found in music, both historical and contemporary, and we discuss its implications for experimental microtonal and spectral music.

Keywords: spectral pitch similarity, harmonicity, affinity, spectrum, melody, microtonality

Testing a Spectral Model of Tonal Affinity with Microtonal Melodies and Inharmonic Spectra

In this paper, we present and experimentally test a psychoacoustic model of the affinity of successive tones in melodies. Based on Terhardt (1984) and Parncutt (1989), we use the term *affinity* to characterize the extent to which successive tones or chords are perceived to have a "good fit", be "unsurprising" or, in some sense, "correct". Affinity is, therefore, a perceptual or cognitive attribute, not a physical attribute; the affinity of two non-simultaneous tones may be thought of as analogous to the consonance of two simultaneous tones. Affinity is important because a preference for certain melodic intervals over others would likely influence "higher-order" musical structures such as scales and chord progressions. For example, we might expect that prevalent scales would contain a preponderance of high-affinity intervals, and that common chord progressions would contain numerous high-affinity intervals between their two sets of tones. Psychoacoustic models of affinity are particularly interesting because they identify sonic features that should be widely perceivable and which operate without prior knowledge of musical structure.

Previous psychoacoustic models of tonal affinity have rested on premises of pitch perception that have not been adequately tested, and have been designed to accommodate only standard Western musical tunings and listeners acculturated to that system. Furthermore, the affinities of successive tones have not been extensively measured prior to this work (two exceptions being Krumhansl 1979 and Parncutt 1989). For these reasons, we have developed a novel psychoacoustic model designed to predict affinities for tones with any spectrum (e.g., harmonic and inharmonic), and intervals of any size (both standard and microtonal). We have also conducted an experiment in which participants ranked the overall affinity of successive tones in melodies played in a variety of musical tunings (some microtonal) and with a variety of tightly controlled spectra (some inharmonic). The resulting model, as optimized to these data, should be applicable to music using both standard tunings and spectra as well as to music with non-standard tunings and spectra.

Background

Most naturally produced sounds (those made by exciting a physical object—banging two rocks together, pushing air through vocal cords, blowing across an open tube, plucking or bowing a taut string, etc.) are *complex tones*, which means they comprise numerous *partials* (frequency components). Furthermore, the sounds produced by most Western musical instruments, including sung vowel sounds, are *harmonic complex tones*, which means that at any given time their partials have frequencies that approximate multiples of a single *fundamental frequency*.

Upon hearing a complex tone, a listener is typically aware of only one or a small number of pitches rather than the full multiplicity of partials physically present. The perceived pitch of a harmonic complex tone typically corresponds to its fundamental, while an inharmonic sound may be heard as comprising more than one pitch (as in a

bell sound), or having a noisy timbre with no identifiable pitch (Moore, 2005; Roederer, 2008). However, all partials that are sufficiently spaced in frequency (greater than the critical bandwidth) are analyzed by the auditory system and can, particularly after training, be individually resolved or "heard out" (brought into awareness) (Helmholtz, 1877; Moore, 2005). For harmonic complex tones, the first nine to eleven partials can usually be heard out (Bernstein & Oxenham, 2003).

Our model encompasses this "complex" nature of sounds by considering the entire spectrum of partials. It also incorporates the uncertainties and inaccuracies of pitch perception resulting from our perceptual (and cognitive) apparatus. The model uses two related components to predict the affinity of a pair of complex tones: (a) *spectral pitch similarity*, which quantifies the *similarity* of tones based on their amplitude spectra (frequencies and amplitudes of partials); (b) *harmonicity*, which quantifies the similarity of the partials of each single complex tone with those of its most similar harmonic complex tone (there are, therefore, two independent harmonicity values for each pair of complex tones). Although neither of these concepts are novel, they have never been run in combination, and our computational formalizations and parameterizations of perceptual uncertainty are original. In the following subsections, we first outline previous research related to spectral pitch similarity, then to harmonicity and, finally, we outline the experimental design.

Overview of pitch similarity models. A set of frequencies (physical phenomena) produce a set of pitches (mental phenomena). A *pitch similarity model* quantifies the perceived similarity of one pitch set (e.g., those resulting from a complex tone or chord) with those of another pitch set (e.g., those resulting from another complex tone or chord).

In the nineteenth century, Helmholtz (1877, Chap. 14) suggested that intervals such as the octave and perfect fifth have a special relationship because, in both cases, so many of the lower tone's partials are replicated in the upper tone. This was extended by Terhardt (1984) who considered not just *spectral pitches*, each of which is evoked by a corresponding partial, but also *virtual pitches*, each of which is evoked by a multiplicity of spectral pitches. A common example of virtual pitch is the way that a harmonic complex tone with a missing fundamental is heard as having a pitch corresponding to that missing fundamental, even though that frequency is physically absent. In Terhardt's (1982) model of pitch perception, a complex tone produces a *profile* of differently weighted spectral and virtual pitches. The precise pitches and weights of the spectral pitches are calculated taking into account auditory masking, thresholds, and sensitivities; the virtual pitches are then generated from these by calculating weighted subharmonics of each spectral pitch and summing them. Terhardt (1984) considered the affinities of tones or chords as arising from them sharing a large number of virtual pitches, rather than sharing a large number spectral pitches.

Parncutt (1989) used Terhardt's pitch model to predict the perceived similarity of successive chords (not tones). This was done by calculating the correlation of their pitch profiles (strictly speaking, this model used a correlation-like function, but this was supplanted by a standard correlation function in Parncutt and Strasburger 1994).

Although not a prerequisite of Parncutt's model, all spectral and virtual pitches were quantized to a 12-tone equal temperament (12-TET) value. The comparative weights of the spectral and virtual pitches were free parameters which, when optimized to the data, strongly favoured the importance of virtual over spectral pitches. Parncutt (1994) later developed a simpler model using only virtual pitches where every notated pitch is assumed to produce a series of candidate virtual pitch classes corresponding to 12-TET-quantized subharmonics. Each such subharmonic has a simple integer weight which approximates the virtual pitch weights produced by Terhardt's model when analyzing a harmonic complex tone with a spectrum typical for a musical instrument (higher partials smoothly decreasing in amplitude). As discussed in Milne, Laney, and Sharp (2015), this model is one of the most effective predictors of Krumhansl's (1982) seminal tonal hierarchies data in which participants rated the fits of all chromatic degrees to a previously established tonal centre (Parncutt, 1994, 2011). Leman's psychoacoustic model also generates virtual pitches (he terms them *periodicity pitches*), and has also been used to model the tonal hierarchies (Leman, 2000) as well as implicit response times to tonal stimuli (Collins, Tillmann, Barrett, Delbé, & Janata, 2014).

In recent work, we have produced similarly effective models of the tonal hierarchies data using only spectral pitches (Milne et al., 2015). Furthermore, with respect to the data collected here, we tried separate spectral pitch and virtual pitch versions of our model and found them to perform similarly well and to be highly correlated (Milne, 2013). Our focus henceforth will be on the spectral pitch model because it is computationally simpler. For our stimuli, virtual pitches provided no advantage; under different stimuli, it may be that including them would be beneficial.

Our model also differs from Parncutt's in a number of other ways. Firstly, we do not quantize our pitches to 12-TET. This quantization assumes listeners are sufficiently acculturated to 12-TET that they cognitively categorize pitches accordingly. We prefer to make no such assumptions so that our model may also apply to listeners familiar with alternative systems (e.g., non-Western or experimental microtonal) as well as to that important period of Western music when harmonic tonality emerged from the earlier modal system. At that time, the chromatic scale was being gradually abstracted out of the prevailing diatonic and hexachordal musical framework and, although we cannot be certain about precisely which tunings were prevalent, it is clear that the musical system was not firmly quantified by precisely 12 categories (thirteenth to fifteenth century treatises present chromatic systems with 12, 14, or 17 tones; Dahlhaus 1990).

Secondly, we assume that the spectral pitch resulting from each frequency component is subject to uncertainty, which is modelled by "smearing" each partial over a range of log-frequencies (as detailed later, this is achieved by convolution with a discrete normal distribution).

Thirdly, we include an additional component also based on spectral content, which is the harmonicity of each complex tone in a pair.

Overview of harmonicity models. As mentioned earlier, harmonicity is a quantification of the similarity of a spectrum to that of the most similar harmonic

complex tone (whose partials are, by definition, multiples of a common fundamental frequency). Although harmonicity is not a model of the relationship between two tones (both are considered separately), it is reasonable to hypothesize that if both tones in a pair are individually heard as in some sense dissonant, complex, unpleasant, or unfamiliar, this will diminish their affinity. This is because listeners cannot separate different aspects of consonance, analogous to many other aspects of perception which tend to be holistic and often multimodal.

An early attempt to demonstrate a link between harmonicity and consonance was made by Stumpf (1890) whose results were subsequently supported by DeWitt and Crowder (1987).[1] Recent experimental results have additionally shown that harmonicity plays an important role in the perceived *pleasantness* of musical chords (McDermott, Lehr, & Oxenham, 2010).

Although harmonicity is widely understood to mean the proximity of a set of partials to those of a harmonic complex tone, few formal mathematical models have been proposed. For example, McDermott et al. (2010) simply use a verbally defined binary harmonic/not harmonic model, which is suitable for their data because they clearly fall into those two categories, but not for less distinct data like ours. Parncutt (1989) provides a method for calculating a possibly related measure called *tonalness*, but this uses the same 12-TET quantization as described above. The MIRtoolbox (Lartillot, Toiviainen, & Eerola, 2008) has an inharmonicity function for a spectrum with I partials, which is $\sum 2\Delta f[i] / I f_{\mathrm{H}}$, where $\Delta f[i]$ is the frequency distance between the spectrum's ith partial and the closest harmonic from a harmonic complex tone with fundamental frequency f_{H}. But we are not aware of any research that has directly tested this, or any other formal mathematical model of harmonicity. For this paper, we develop our own model of harmonicity (detailed in the next section), which utilizes the same underlying methods as our model of spectral pitch similarity. An advantage of this is that spectral pitch similarity and harmonicity can be expressed in the same units and hence are made directly comparable.

In our experiment, we use a set of spectra with differing harmonicities. All of our spectra also had fairly widely spaced partials so they were all relatively smooth (they did not exhibit audible beating). For these stimuli, therefore, a roughness model (e.g., Kameoka and Kuriyagawa 1969; Plomp and Levelt 1965; Sethares 2005) would be superfluous.

Overview of experiment. Our experiment was specifically designed to test our overall model of affinity, as well as the individual impacts of spectral pitch similarity and harmonicity. Forty-four participants listened to melodies played in a variety of equal tunings: in addition to the familiar 12-TET, which divides the octave into twelve equal parts (frequency ratios), we used an additional ten different equal divisions of the octave, most of them producing microtonal intervals not found in 12-TET. The full list of tunings used is 3-TET, 4-TET, 5-TET, 7-TET, 10-TET, 11-TET, 12-TET, 13-TET, 15-TET, 16-TET, 17-TET (all intervals in 3- and 4-TET are also found in 12-TET; all other n-TETs in this list produce intervals not found in 12-TET). Melodies

[1] The subtleties of Stumpf's theories of consonance are discussed in Schneider (1997).

were used as stimuli, rather than isolated intervals, in order to more closely reflect the way that real-world music is heard and assessed.

Given an *n*-TET, each melody was randomly generated with a probability distribution, over note transitions, designed to model common features of melodies; for example, making small steps more common than large leaps. This was done to minimize distraction and maximize ecological validity. Random generation was used to avoid any unintentional bias towards stimuli supporting our hypotheses that may have arisen if the melodies had have been composed by ourselves. For each melody, the tempo and articulation (tone duration as a percentage of interonset interval) was also randomly chosen (within an overall range of values that would be common in musical performance). A large number of melodies were tested (2638) to ensure any additional effects induced by the randomly chosen tempos, articulations, contours, actual pitch choices, and so forth, had minimal bias on our variables of interest (spectral pitch similarity and harmonicity).

Each such melody was played with two different spectra, and participants chose which of the two "timbres" produced the greater overall affinity. A binary forced-choice was used (rather than individually rating each melody) to ensure the task was both simple to perform whilst still being sensitive to possibly small effects. One of the spectra was *matched* to the melody's *n*-TET to ensure the average spectral pitch similarity between successive tones was relatively high, while the other spectrum was *unmatched* and so its average spectral pitch similarity between successive tones was lower. For expediency, the spectral matching was achieved with existing software (the synthesizer The Viking; Milne and Prechtl 2008). The spectral matching method used by this synthesizer is detailed in Sethares, Milne, Tiedje, Prechtl, and Plamondon (2009) and outlined further in the "Methods" section. But, in brief, all partials in the sound are tuned to a frequency found in the *n*-TET to which it is matched. The tunings of the lowest twelve partials, when matched to each of the above eleven *n*-TETs, are shown in Table 1.

The matched and unmatched spectra also had different levels of harmonicity because some *n*-TETs allow closer approximations of the frequencies in a harmonic complex tone than do others. All 110 different pairs of matched and unmatched spectra were tested; for example, there was a stimulus with a 5-TET melody played with a matched spectrum in 5-TET and an unmatched spectrum in 11-TET as well as a *complementary* stimulus with an 11-TET melody played with an 11-TET matched spectrum and a 5-TET unmatched spectrum.

Having complementary pairs ensures that: (a) any overall preference for matched spectra cannot be down to harmonicity; (b) spectral pitch similarity and harmonicity were uncorrelated across the differing melodies (as confirmed in the "Results" section), which enables the influence of these two components to be disambiguated. Having the same melody for the matched and unmatched spectra in each forced choice ensures that (c) interval size played no role in participants' choices because the interval sequence was always the same for the two versions of the melody, which removes an important long-term memory confound. Together, these imply that an overall preference for matched spectra (higher spectral pitch similarity)

Table 1
*The log-frequencies (relative to the first partial and rounded to the nearest cent) of the partials of a harmonic complex tone (*HCT*) and the spectra matched to the n-*TETs *used in the experiment.*

	Partial number											
Spectrum	1	2	3	4	5	6	7	8	9	10	11	12
HCT	0	1200	1902	2400	2786	3102	3369	3600	3804	3986	4151	4302
3-TET	0	1200	2000	2400	2800	3200	3600	3600	4000	4000	4000	4400
4-TET	0	1200	1800	2400	2700	3000	3300	3600	3600	3900	4200	4200
5-TET	0	1200	1920	2400	2880	3120	3600	3600	3840	4080	3840	4320
7-TET	0	1200	1886	2400	2743	3086	3257	3600	3771	3943	4286	4286
10-TET	0	1200	1800	2400	2760	3000	3120	3600	3600	3960	4320	4200
11-TET	0	1200	1964	2400	2836	3164	3382	3600	3927	4036	4145	4364
12-TET	0	1200	1900	2400	2800	3100	3400	3600	3800	4000	4100	4300
13-TET	0	1200	1846	2400	2769	3046	3231	3600	3692	3969	4246	4246
15-TET	0	1200	2000	2400	2800	3200	3600	3600	4000	4000	4000	4400
16-TET	0	1200	1875	2400	2775	3075	3300	3600	3750	3975	4200	4275
17-TET	0	1200	1906	2400	2824	3106	3459	3600	3812	4024	4024	4306

cannot be influenced by long-term statistical learning of the prevalences of differing interval sizes or differing harmonicities. Any overall preference for high-harmonicity tones may, however, be due to long-term statistical learning.

In summary, we use our model and data to test three principal hypotheses: (a) affinity is a monotonically increasing function of spectral pitch similarity; (b) affinity is a monotonically increasing function of harmonicity;[2] (c) spectral pitch similarity is modelling a psychoacoustic process that operates even in the absence of prior learning of interval prevalences. Given the experimental design, which eliminates any impact of interval familiarity, evidence for the last hypothesis follows directly from evidence for the first.

The Models

The respective purpose of each model is to numerically quantify the spectral pitch similarity of any two sounds and to numerically quantify the harmonicity of any single sound. The additional methods we used to apply these models specifically to our experimental data (which comprise binary choices made with respect to complete melodies rather than single intervals) are given in the "Results" section. As described above, these two variables are then used to model affinity. We will consider harmonicity to be the spectral pitch similarity of a sound with its most similar harmonic complex tone, which can be thought of as a "template" (the precise form of this template will be discussed later).

Both models, therefore, require a mathematical formalization of spectral pitch similarity. At the outset, it is useful to state that there is no single most simple, canonical, or "natural" measure of the similarity of two spectra. For example, it may seem straightforward to total up the log-frequency distances between pairs of partials

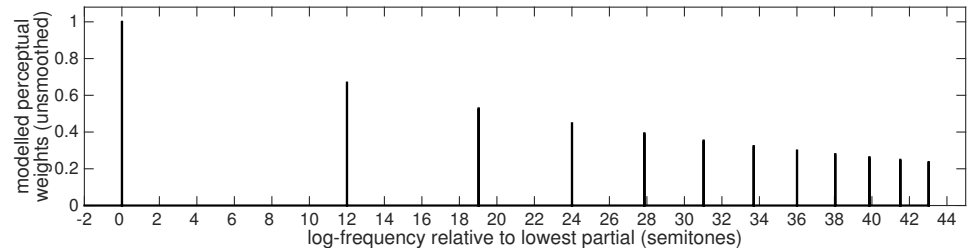
[2] *Monotonically increasing* means that affinity does not get smaller when the predictor's value increases (all else being equal). The precise relationship between them may, however, be non-linear; e.g., it might approximate a power function like many psychophysical variables (Stevens, 1957).

(each pair containing one partial from each tone), but this method would be restricted in scope because it is applicable only to tones with identical numbers of partials. Furthermore, it is not obvious why each partial in one tone should be uniquely paired with single partial in the other tone, and precisely how those pairings should be chosen.[3] This approach is also founded on the unlikely presumption that the perceptual system is able to independently track the "motions" of numerous simultaneously sounding partials.

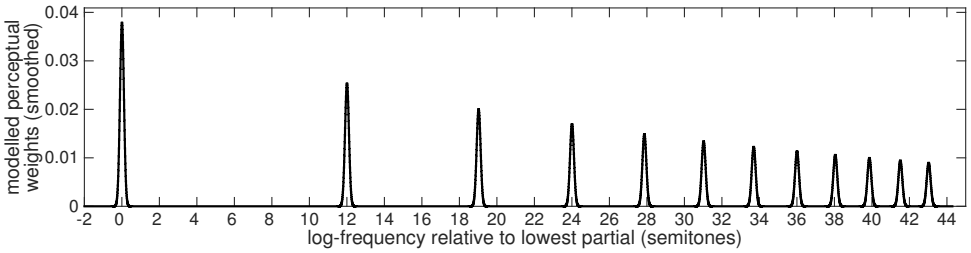
A more generally applicable and perceptually plausible approach, now described, is to consider the proportion of partials in the two tones that correspond in pitch (under reasonable expectations of perceptual pitch uncertainty).

Spectral Pitch Vectors

The models for both spectral pitch similarity and harmonicity are based on the expectation tensors introduced in Milne et al. (2011). In this case, the tensor is of the simplest kind—a *spectral pitch vector* in which delta spikes, which indicate the log-frequencies (in cents) and perceptual weights of all partials, are *smoothed* with a discrete normal distribution. This is illustrated in Figure 1.



(a) Unsmoothed spectrum.



(b) Smoothed spectrum.

Figure 1. Spectral pitch vectors showing the effect of smoothing (convolving) a set of harmonic partials with a discrete approximation of a normal distribution with a standard deviation of $\sigma = 10.53$ cents. The roll-off is $\rho = 0.58$. These are the parameter values as optimized to the experimental data, as detailed later. The weights on the vertical axis model the expected numbers of partials perceived within each log-frequency bin in the vector.

[3] An apparent solution would be to pair partials that are closest in log-frequency but, even ignoring potential ambiguities, the resulting function would no longer be a true distance metric. The same problem also applies to the MIRtoolbox method described earlier. These issues are discussed in depth in Milne, Sethares, Laney, and Sharp (2011).

The width of the smoothing is a free parameter (σ, the Greek letter *sigma*), and the steepness of the roll-off in the weighting of ascending harmonics is another free parameter (ρ, the Greek letter *rho*). The smoothing-width parameter models the perceptual inaccuracies that result in close, but non-identical, frequencies being judged as having the same pitch—the greater the width of the normal distribution the greater the modelled perceptual inaccuracy. The roll-off parameter models the lesser perceptual importance of higher partials relative to lower partials. This will likely depend on the spectrum used for the stimulus, but this parameter additionally allows the model to take account of psychoacoustic processes. For example, it is easier to perceptually resolve (consciously hear out) lower harmonics than it is higher harmonics, even when they have equal intensity (Bernstein & Oxenham, 2003; Moore, 2005).

More formally, for any given tone, a many-element row-vector of zeros is created (typically there will be thousands of elements). The first element represents the log-frequency of the lowest partial under consideration. The second element is one cent higher, the third element is two cents higher, and so forth. The vector needs to have a sufficient number of elements to ensure the last is at least as high in log-frequency as the highest partial under consideration. For each of the partials in the tone, a value of unity is placed in the element corresponding to its log-frequency (cents) value, all other entries are zero. These values are denoted *weights*.[4] We additionally index each partial by i, such that $i = 1$ is the lowest partial, $i = 2$ is the next higher partial, and so forth.

To apply the roll-off, we multiply the weights by $1/i^{\rho}$. When $\rho > 0$, this means every higher partial has a lesser weight than every lower partial, but no partial has a negative weight. The steepness of the roll-off is determined by the size of ρ. An example of the type of vector that results is illustrated in Figure 1(a). To apply the smoothing, we convolve the ρ-weighted vector with a discrete normal distribution with a standard deviation of σ. The effect of this smoothing is illustrated in Figure 1(b). The resulting vector is denoted a *spectral pitch vector*. We use the term *pitch*, rather than *log-frequency* or *cents*, because the smoothing and weights are modelling perceptual processes that have "transformed" the original acoustical stimulus.

For our analysis, we include only the first twelve partials in the spectral pitch vectors. This is because partials higher than this typically cannot be perceptually resolved (Bernstein & Oxenham, 2003), and removing them from the model reduces the number of calculations required (the computational efficiency of the model becomes a major concern under optimization to the data, particularly when cross-validating). Given that we do not include partials higher than the twelfth, we would expect the optimized value of ρ to approximately correspond to the

[4]More formally, we can consider each weight as a model of the probability of that partial being perceived. This then implies that the values in the resulting spectral pitch vector (described subsequently) model the expected number of partials perceived at each log-frequency bin in the vector, and that the spectral pitch similarity (cosine similarity) of any two such vectors is equivalent to the proportion of partials in the two tones that correspond in pitch (Milne et al., 2011).

loudnesses of the partials in the sonic stimuli actually used. As discussed in Milne et al. (2011, App. A, Online Supplementary), the smoothing width σ models the just noticeable frequency difference, which is 3–13 cents between 125 and 6000 Hz (Moore, 1973). We would, therefore, expect the optimized value of σ to be within or close to this range of cents values. Both these expectations were subsequently confirmed by the data, as shown in the "Results" section.

The spectral pitch vectors described here are a relatively simple model in that they do not take into account the additional complexities of pitch perception embodied in Terhardt's (1982) model (e.g., frequency and amplitude masking), and because they only approximate the actual signal with the ρ parameter. However, our purpose here is to ascertain in a general way whether spectral pitches play a perceptually meaningful role in a variety of melodic stimuli. In future research, it might be interesting to compare our model with one that takes into account these additional effects.

Spectral Pitch Similarity Model

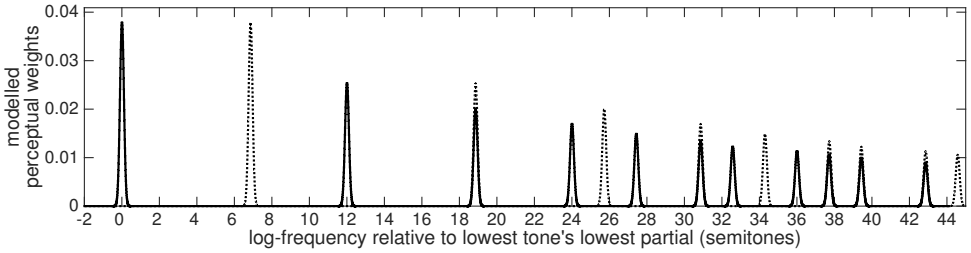
The *spectral pitch similarity* of any two tones is simply modelled as the *cosine similarity* between their respective spectral pitch vectors. Cosine similarity is the cosine of the angle between the two vectors.[5] For vectors all of whose values are positive (as is the case for spectral pitch vectors), their cosine similarity is always between zero (maximally dissimilar) and unity (maximally similar). The cosine similarity of two spectral pitch vectors (both row vectors) denoted $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by $s(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}\boldsymbol{y}^\mathsf{T} / \sqrt{\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\, \boldsymbol{y}\boldsymbol{y}^\mathsf{T}}$, where $^\mathsf{T}$ is the transpose operator that converts a row vector into a column vector, and the multiplications are all matrix multiplications.

Figure 2 illustrates two pairs of spectral pitch vectors. The first (Fig. 2(a)) is from a pair of tones a 7-TET "fifth" (of 6.86 semitones) apart, the lower-pitched tone drawn with a solid line, the higher-pitched tone with a dotted line. Both spectra are matched to 7-TET, hence the spectrum matches the tuning. Note how their partials' frequencies perfectly coincide at numerous log-frequencies. The second (Fig. 2(b)) is a pair of tones the same interval apart, but now they have spectra that are unmatched—they are matched to 11-TET. Note how the partials no longer coincide—the only location where their distributions overlap is around 41 semitones. This visualizes how the first two spectra are more similar than are the second two; the cosine similarity values given in the captions precisely quantify this.

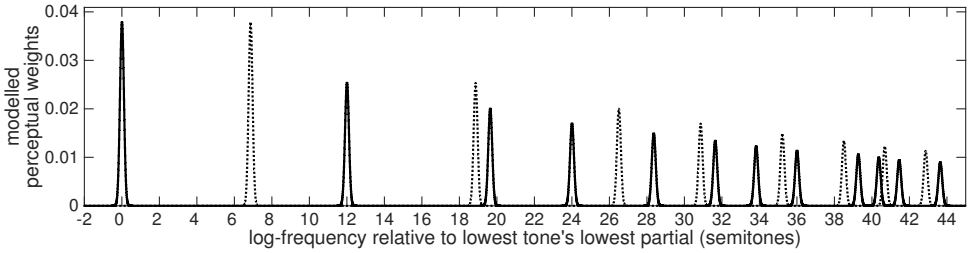
Harmonicity Model

The *harmonicity* of a tone is modelled by calculating the spectral pitch similarity of a spectral pitch vector and a "template" harmonic complex tone's spectral pitch vector, over all possible cents transpositions of the latter. This can be thought of as a normalized cross-correlation of the two vectors. The maximum value is then extracted from the resulting vector and this serves as the harmonicity value. (This is

[5] An advantage of cosine similarity over an L_p metric like Euclidean and Manhattan is that, for non-negative vectors like spectral pitch vectors, its value is conveniently bounded between 0 and 1, and is dimensionless (like a correlation value, it is unaffected by the units in which its arguments are measured). A further advantage is that—as described earlier—its meaning in this context is easy to interpret: it models the proportion of partials in the two tones that correspond in pitch (given ρ and σ).



(a) Two tones in a 7-TET tuning with matched spectra. Their spectral pitch similarity is .294.



(b) Two tones in a 7-TET tuning with unmatched spectra (they are matched to 11-TET). Their spectral pitch similarity is .003.

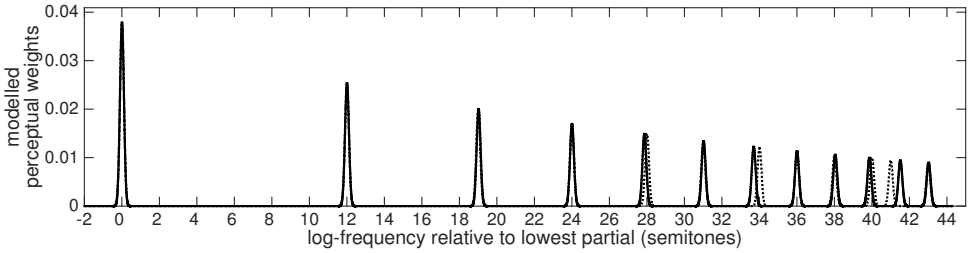
Figure 2. Both figures show the spectral pitch vectors for a pair of tones a 7-TET "fifth" of 6.86 semitones apart. The lower-pitched tone in each pair is drawn with a solid line, the higher-pitched tone with a dotted line. In (a), the spectra are matched to the tuning, in (b) they are not (the spectra are matched to 11-TET). The spectral pitch similarity values are calculated under the model's optimized parameter values.

related to the approach introduced by Brown 1992, which uses cross-correlation, in the log-frequency domain, of a complex tone and a harmonic complex template to estimate the former's fundamental.) For the sake of simplicity and parsimony, the roll-offs and smoothing widths of both the template and the tone are determined by the same ρ and σ values as used for the spectral pitch similarity model. This also means that, regardless of the value of ρ, if a spectrum's partials perfectly coincide in frequency with those of the template, it will have the maximum possible harmonicity of 1 (this would not be the case if the template had fixed spectral weights, e.g., every partial has a weight of unity as in Brown's (1992) model).

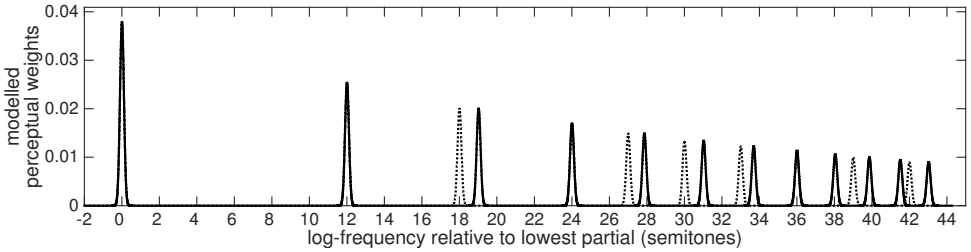
Figure 3 illustrates two pairs of spectral pitch vectors. In each case, one spectrum is from a complex harmonic tone, the other is one that has been matched to 12-TET (a) or 4-TET (b). They illustrate how the first pair are more similar (they have greater overlap) than the second pair, hence the 12-TET spectrum has higher harmonicity than the 4-TET spectrum. This harmonicity (similarity with a harmonic template) is precisely quantified by the cosine similarity values in the captions.

Experimental Method

This section begins with a description of how the sounds were synthesized, before moving on to discuss the melody generation. Audio examples are available from the supplemental online section. After that, the delivery of the experiment is described.



(a) A spectrum matched to 12-TET (dotted line) and the most similar harmonic complex spectrum (solid line). Their spectral pitch similarity, and hence the harmonicity of the 12-TET spectrum, is .910.



(b) A spectrum matched to 4-TET (dotted line) and the most similar harmonic complex spectrum (solid line). Their spectral pitch similarity, and hence the harmonicity of the 4-TET spectrum, is .669.

Figure 3. Both figures show the spectral pitch vectors for a harmonic complex tone (solid line) and a spectrum matched to an *n*-TET (dotted). In (a), the latter is matched to 12-TET; in (b), it is matched to 4-TET. The spectral pitch similarity values are calculated under the model's optimized parameter values.

Spectral Matching

The method used to match the spectrum to an *n*-TET scale is fully detailed in the "Dynamic Tonality" section of Sethares et al. (2009). In summary, the log-frequency of each partial is an *n*-TET approximation of what it would be in a harmonic complex tone. Clearly some *n*-TETs will provide better approximations than others, so the harmonicities of spectra matched to differing *n*-TETs have considerable variance. Because the intervals between the harmonics are from the same *n*-TET as the underlying tuning, successive tones typically have one or more partials with identical log-frequencies. This implies that the intervals in melodies using matched spectra will typically have greater spectral pitch similarity than when using unmatched spectra. Furthermore, because all deviations from harmonicity are by log-frequency, all interval sizes between successive tones (which are also measured in terms of log-frequency) are unchanged by different spectral tunings (the log-frequencies of the partials for all matched spectra were summarized earlier in Table 1). This specific selection of *n*-TETs was chosen for expediency: they were the eleven lowest values of *n* supported by the spectral-matching synthesizer we used to generate the melodies.

The amplitude (power) of each partial was $1/i$ where i is the number of the partial (if all partials had been in the same phase and tuned to a harmonic series this would give a sawtooth waveform). Each tone was enveloped with a quick, but non-percussive, attack and a full sustain level. With harmonic partials, the timbre

sounded somewhat like a brass or bowed-string instrument. To slightly mellow the sound, the tones were then passed through the synthesizer's low-pass filter set to give a small resonant peak. The filtering had only a minor impact on the magnitudes of the partials. A small amount of delayed-onset vibrato was added to give the sound life, and a small amount of reverberation/ambience to emulate the sound of a small recital room.

Melody Generation

Every melody contained 16 eighth notes (e.g., two bars of $\frac{4}{4}$, although there was no rhythmic accentuation to imply any specific meter). The melody was randomly generated (hence different) for each presentation of a matched and unmatched pair of spectra (though identical for each such pair). We constructed a parameterized probability distribution specifying the probabilities for all note transitions. This distribution was designed to emulate common features of melodies (both Western and non-Western) so as to avoid distracting the participants with unfamiliar melodic constructions (beyond the unfamiliarity engendered by the microtonal tunings), and to allow the results to generalize better to real-world music. Precisely the same parameter values were applied to all eleven tunings used in the experiment.

We now outline the musical features we emulated: (a) in Western and non-Western melodies, smaller intervals typically occur more often than larger intervals (Vos & Troost, 1989, and references therein); (b) the average notated pitch of both Western and non-Western music is approximately D♯4 (three semitones above middle C) (Parncutt 1992, cited by Huron 2001); (c) conventional Western melodies principally comprise pitches from pentatonic or diatonic scales—although chromatic pitches do occur, they are less common; (d) modulations (scale transpositions) are infrequent. The methods used to generalize these to microtonal tunings, and the precise modelling and parametrization used to do this, are provided in Appendix A.

For each different melody the interonset interval for eighth-notes was randomly chosen, with a uniform distribution, over the range 163–476 ms (63–184 beats per minute), whose mean of 319.5 ms (94 bpm) equates to a medium tempo; the articulation (ratio of note-length to interonset interval) was randomly chosen from the range 0.72 to 0.99, whose mean of 0.86 equates to the average articulation used by organists (Jerket, 2004).

Participants

Forty-four academic and non-academic university staff and graduate students participated in the experiment (25 male, 19 female, mean age 37.4 years, standard deviation 11.1 years), and no reimbursement was given. Eleven reported to have had no musical training or ability; 12 to have had basic musical training or ability (Associated Board of the Royal Schools of Music Grades 1–4, or similar qualification or experience); 14 to have had intermediate training or ability (Grades 5–7, or similar); 7 to have had advanced training (Grade 8 or higher, or similar). The average level is, therefore, somewhere between basic and intermediate, and the overall distribution is wide. None claimed to possess absolute pitch ("perfect pitch").

Forty-four participants were chosen in order to ensure each stimulus (as characterized by its matched and unmatched timbral tunings) was tested by a number of participants sufficiently large to detect small-sized effects and to ensure a broad range of participants took part (as characterized by musical experience, taste, age, etc.). Due to the experimental design, each such stimulus was rated by an average of twenty-four participants (the precise numbers are given in Table B1).

Apparatus

The tones were generated by a modified version of The Viking v1.0 (Milne & Prechtl, 2008), which is a freeware additive-subtractive synthesizer built within Outsim's SynthMaker with the capacity to match spectrum and tuning. The synthesizer's tuning parameters and notes were controlled by live MIDI generated by a patch written in Cycling 74's Max/MSP. The patch used the random probability distributions specified earlier. The patch (and accompanying JavaScript routine), and the modified version of The Viking can be downloaded from the online supplemental section.[6] The stimuli were played over closed-back headphones (Audio Technica ATH-M40fs) in a quiet room.

Procedure

Each participant listened to 60 different randomly generated melodies. Each melody was played in an n-TET randomly chosen from eleven possibilities: 3-TET, 4-TET, 5-TET, 7-TET, 10-TET, 11-TET, 12-TET, 13-TET, 15-TET, 16-TET, and 17-TET. For each melody, the participant could use a mouse or touchpad to select between two vertically arranged radio buttons. Each button produced a different spectrum: one spectrum was *matched* (its partials were in the same n-TET tuning as the melody); the other was *unmatched* (its partials were in an n-TET tuning different to the melody, randomly chosen from the same list). Each melody could be repeated, by the participant, as many times as wished. The buttons to which the matched and unmatched spectra were mapped were randomly chosen for each melody. No mention was made to the participant that the buttons changed the spectrum or timbre. For each melody, the participant was asked to indicate the button where the different notes of the melody had the greatest affinity, which was clarified by the following criteria: they have the greatest affinity; they fit together the best; they sound most in-tune with each other; they sound the least surprising. These four descriptions constitute our operationalization of affinity. All participants claimed to understand the task prior to starting.

Most trials were completed in 25–30 minutes. For each participant, no pair of underlying tuning and unmatched spectral tuning occurred more than once. There are 110 different possible stimuli (pairs of distinct matched and unmatched spectra). The 60 different stimuli listened to by each participant were sampled randomly without replacement from the 110. This means that, on average, each stimulus has been tested $44 \times 60/110 = 24$ times, each underlying tuning (and associated matched spectrum) $44 \times 60/10 = 264$ times, each unmatched spectral tuning $44 \times 60/10 = 264$

[6] The current publicly available version of The Viking (`http://www.dynamictonality.com`) has since been completely recoded in Max/MSP after the experiment was conducted.

times. In total there were $44 \times 60 = 2640$ observations of 110 different stimuli. Two tests were lost due to the experiment ending prematurely, giving a total of 2638 tests.

Results

In the first subsection, we provide some straightforward analyses of the experimental data without recourse to our models of harmonicity and spectral pitch similarity. In the second, we explain how our models of spectral pitch similarity and harmonicity are applied to these data, and we explore whether they can more comprehensively explain the data, notably by separating out the individual impacts of spectral pitch similarity and harmonicity. The raw data can be downloaded from the supplementary online section.

Data Analysis

Our first hypothesis was that affinity is a monotonically increasing function of spectral pitch similarity. If true, we would expect participants to choose matched spectra more often than unmatched. Of the 2638 tests, matched spectra were chosen 1615 times (61% of occasions, with a 95% binomial confidence interval from 59% to 63%). Given the null hypothesis that the use of matched or unmatched spectra has no influence on melodic affinity, the expected number of matched spectra chosen would be $.5 \times 2638 = 1319$ with a binomial distribution of $\text{Bin}(2638, .5)$. Under this null hypothesis, a two-tailed exact binomial test shows the probability of 1615, or greater, matched spectra being chosen is $p < .001$ (the actual p-value is smaller than the level of computational precision and is reported by MATLAB as zero). Indeed, 1370 (52%) is the minimum number of matched spectrum choices that would have been significant at the .05 level. This supports our first hypothesis.

Of the 44 participants, 38 (86%) chose matched spectra for more than half of the 60 stimuli they listened to. Under the null hypothesis that 50% of participants would choose matched spectra more often than unmatched, an exact binomial test (two-tailed) shows the probability of this occurring by chance is $p < .001$. This indicates preference for matched spectra was not confined to a small number of "high performing" participants, thereby providing further evidence in support of the first hypothesis, and its generality across different individuals.

Our second hypothesis was that affinity is a monotonically increasing function of harmonicity. This requires a more detailed analysis and visualization of the data. The data for all 110 different stimulus pairs (matched and unmatched spectra), aggregated over all participants, are summarized in Figure 4 (the same data are also summarized in tabular form in Appendix B). The shade of each square indicates the ratio of occasions when the matched spectrum was chosen rather than the unmatched—white would be 100% matched, black would be 0% matched. (Henceforth, we use the terminology "ratio of matched spectra chosen", or similar, to mean the number of matched spectra chosen divided by number of matched and unmatched spectra chosen, for the group of stimuli under consideration.) The vertical axis shows the n-TET used for the underlying tuning (equivalently, the tuning of the matched spectrum's partials); the horizontal axis shows the n-TET used for the tuning of the unmatched spectrum's partials. For example, the square on the row

marked 7 and the column marked 11 shows the ratio of occasions that, for a 7-TET melody, the matched spectrum (partials tuned to 7-TET) was chosen rather than the unmatched spectrum (partials tuned to 11-TET).

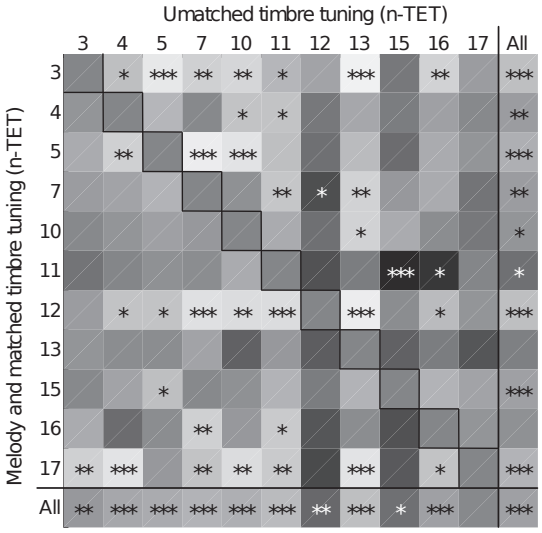
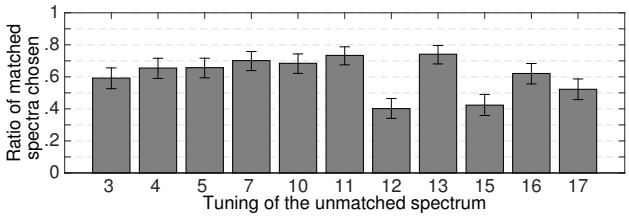


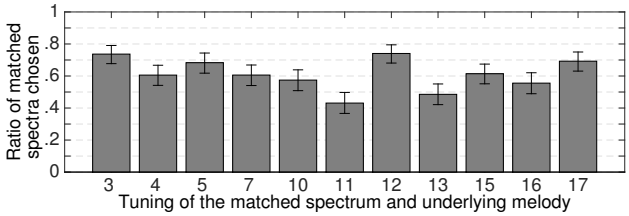
Figure 4. Results aggregated over all participants. The shade indicates the ratio of matched timbres chosen (white = 100%, black = 0%) for each tested pair of matched and unmatched spectra. Stars indicate significance levels—black for higher than the null hypothesis, white for lower (Bonferroni correction has not been applied—see the main text).

The squares in the top-left to bottom-right diagonal (they have thicker borders) would correspond to situations where both spectra are identical. Such pairs were not tested because it is clear that—given the forced-choice nature of the procedure—the probability of choosing either would converge to .5. For this reason, the diagonal is shaded accordingly, and this serves as a useful reference against which to compare the other data points. The bottom row shows the ratios of matched spectra chosen, aggregated over all possible tunings, for each of the eleven unmatched spectra (this is also shown in Fig. 5a). The rightmost column shows the ratio of occasions a matched spectrum was chosen, aggregated over all possible unmatched spectra, for each of the eleven underlying tunings (this is also shown in Fig. 5b). The bottom-right square shows the ratio of occasions a matched spectrum was chosen aggregated over all underlying tunings and unmatched spectra (the previously discussed ratio of 61%).

A single star indicates a ratio that is significantly different from .5 (using a two-tailed exact binomial test) at a level of .05, two stars indicate significance at the .01 level, three stars at the .001 level. We have not applied Bonferroni correction here, because we are not inferring a preference for matched partials on the basis of any single stimulus, and it is interesting to see which of the stimuli are sufficiently different from chance to merit individual significance. It is worth noting that with 110 separate tests we would expect 5.5 to be significant at the .05 level under the null hypothesis of pure chance (2.75 higher, 2.75 lower). In actuality, there are 32 stimuli where the matched spectrum was chosen significantly more often than expected



(a) Over different unmatched spectra—the bottom row of Fig. 4.



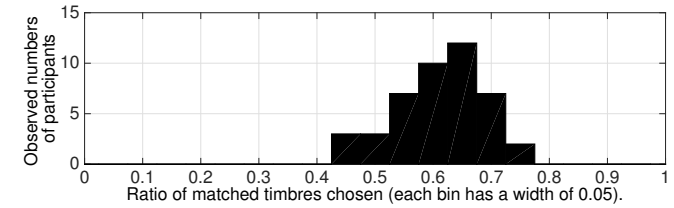
(b) Over different matched spectra (underlying tunings)—the rightmost column of Fig. 4.

Figure 5. Ratios of matched spectra chosen (i.e., the number of times a matched spectrum was chosen divided by the number of times a matched or unmatched spectrum was chosen). The error bars show the 95% binomial confidence intervals (as calculated by the Clopper-Pearson method).

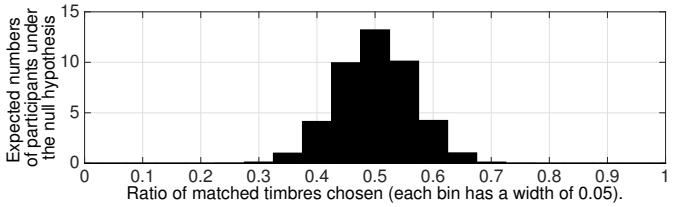
under the null hypothesis, and 3 stimuli where the matched spectrum was chosen significantly less often than expected under the null hypothesis.

Figure 4 illustrates some interesting vertical and horizontal stripes. For example, the columns representing the unmatched spectra tuned to 12-TET, 15-TET and 17-TET are darker; indeed, their aggregated probabilities (as shown by the bottom row and Fig. 5(a)) are all significantly lower than the overall mean probability of 61%. This indicates that participants felt these spectra tended to have relatively higher affinity regardless of the underlying tuning. This is interesting because these three spectra all have partials that are relatively close to perfectly harmonic partials (our harmonicity model subsequently confirms this). The horizontal stripes, which represent the underlying tuning and its matched spectrum, are complementary to the vertical stripes. For example, if the 12-TET spectrum is preferred regardless of tuning then, when the underlying tuning—and its matched spectrum—is 12-TET, the unmatched spectra are now less likely to be chosen. Hence, the corresponding row is lighter. So the dark vertical stripes and corresponding light horizontal stripes are complementary manifestations of the same process. These vertical and horizontal stripes, therefore, support the hypothesis that affinity is a monotonic function of harmonicity.

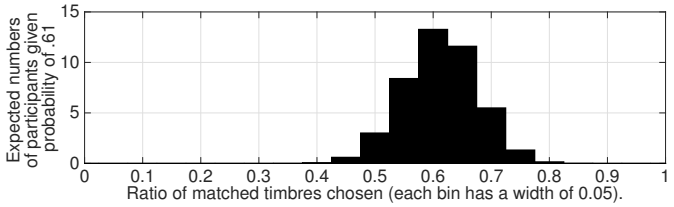
Figure 6(a) is a histogram showing the distribution of participants' responses, binned according to the overall ratio of matched spectra chosen. The bins have a width of 0.05 (following Sturge's (1926) rule). Figure 6(b) shows the histogram that would be expected under the null hypothesis of a uniform .5 probability of choosing a matched spectrum. Figure 6(c) shows the histogram that would be expected under



(a) As observed in the experiment.



(b) As expected under a uniform probability of .5.



(c) As expected under a uniform probability of .61.

Figure 6. The observed histogram of participants' ratios of matched spectra chosen over all stimuli (a). For comparison, (b) and (c) are the expected histograms arising from two different hypotheses. Their values are the means of multiple histograms randomly generated under $p = .5$ and $p = .61$ over all participants and all stimuli.

a hypothesis of a uniform .61 probability (the observed mean probability) of choosing a matched spectrum.

The histograms indicate that participants' responses were consistent. This is demonstrated by the similarity of the shape and range of the observed histogram (Fig. 6(a)) with both of the others (Fig. 6(b) and (c)). Binomial distributions are negatively skewed (the left tail is longer) when the mean is $> .5$, as can be seen in Figure 6(c). In comparison with the .61 histogram, the observed histogram has a slightly heavy left-hand tail. This may indicate the presence of a small number of participants for whom the impact of spectral pitch similarity was negligible.

Data Modelling

Our models for the spectral pitch similarity of successive tones and for the harmonicity of individual tones were specified in the earlier "Models" section. In this section we describe how these models are applied to our data, which comprises binary choices made in response to complete melodies. We subsequently refer to these two models as "predictors", because they are both used as such in a full model—a multiple logistic regression—which is fitted to the data. We evaluate the model's goodness of fit to the data and analyse the implications of the optimized parameter values. The full model, and its optimization and analysis, can be downloaded as MATLAB routines from the supplementary online section.

The spectral pitch similarity predictor. Participants' rankings of affinity were based on the melodies as a whole (one with a matched spectrum, the other with an unmatched spectrum). Each melody had 16 notes, hence 15 intervals between successive tones, so we needed a way to model their overall impact. An obvious and simple way to do that is to take the mean spectral pitch similarity across all 15 intervals. However, each melody was randomly generated and to make this calculation for all 2638 distinct presentations would make optimization and cross-validation of the resulting model prohibitively slow. Instead, we estimated this by calculating the expected number of occasions each interval would occur—in each of the eleven distinct n-TETs—due to the stochastic process used to generate them (the full parameterization of this stochastic process is detailed later). These expected numbers of intervals were then used to calculate an estimate of the *expected spectral pitch similarity*.

The final *spectral pitch similarity predictor* is the expected pitch similarity of the matched spectrum minus the expected pitch similarity of the unmatched spectrum. Given a melody, played with a given matched and unmatched spectra, this predictor is used to model the probability of participants choosing the matched spectrum as producing the greatest overall affinity. We use the mathematical notation $f_{\Delta S}(\boldsymbol{r}_{n_1}, \boldsymbol{s}_{n_1}, \boldsymbol{s}_{n_2}; \rho, \sigma)$, which means it is a function of the independent variables $\boldsymbol{r}_{n_1}$ (the expected numbers of all different interval of differing sizes given the melody's n-TET tuning), $\boldsymbol{s}_{n_1}$ (the matched spectrum, which is tuned to the same n-TET as the melody) and $\boldsymbol{s}_{n_2}$ (the unmatched spectrum, which is tuned to a different n-TET), and it is parameterized by ρ (roll-off) and σ (smoothing width).

The harmonicity predictor. The *harmonicity predictor* is simply the harmonicity of the matched spectrum minus the harmonicity of the unmatched spectrum. Given matched and unmatched spectra, this predictor is used to model the probability of participants choosing the matched spectrum as producing the greatest overall melodic affinity. Unlike the spectral pitch similarity predictor, this predictor is not dependent on the actual melody used, only the two spectra. We use the mathematical notation $f_{\Delta H}(\boldsymbol{s}_{n_1}, \boldsymbol{s}_{n_2}; \rho, \sigma)$, which means it is a function of the independent variables $\boldsymbol{s}_{n_1}$ (the matched spectrum tuned to the same n-TET as the melody) and $\boldsymbol{s}_{n_2}$ (the unmatched spectrum, which is tuned to a different n-TET), and it is parameterized by ρ (roll-off) and σ (smoothing width), which are the same as those used in the spectral pitch similarity model.

The full model. For each presentation of a melody, a datum Y was coded 1 when the matched spectrum was chosen and coded 0 when the unmatched spectrum was chosen. The data were aggregated for each pair of matched and unmatched spectral tuning in order to estimate the probability of choosing the matched spectrum for that pair of spectral tunings. Consider a hypothetical example where 24 participants hear a melody in 11-TET (so the matched spectrum is also tuned to 11-TET), and an unmatched spectrum tuned to 7-TET; of these participants, 13 choose the matched spectrum as having greater affinity, and 11 choose the unmatched as having greater affinity. This means the estimated probability of choosing a matched spectrum, under these spectral tunings, would be 13/24. In a logistic regression,

these probabilities are modelled accordingly:

$$P\big(Y = 1 \mid \boldsymbol{r}_{n_1}, \boldsymbol{s}_{n_1}, \boldsymbol{s}_{n_2}\big) = \frac{1}{1 + \mathrm{e}^{-z}}\,, \text{ where}$$
$$z = \beta_1\, f_{\Delta\mathrm{S}}(\boldsymbol{r}_{n_1}, \boldsymbol{s}_{n_1}, \boldsymbol{s}_{n_2}; \rho, \sigma) + \beta_2\, f_{\Delta\mathrm{H}}(\boldsymbol{s}_{n_1}, \boldsymbol{s}_{n_2}; \rho, \sigma)\,. \qquad (1)$$

The logistic regression weight (the β_1 coefficient) applied to the spectral pitch similarity predictor, the logistic regression weight (the β_2 coefficient) applied to the harmonicity predictor, and the nonlinear roll-off (ρ) and smoothing (σ) parameters were all optimized simultaneously to maximize the likelihood of the model given the data.

If the "matched" and "unmatched" spectra were to be identical (which implies $f_{\Delta\mathrm{S}} = f_{\Delta\mathrm{H}} = 0$), participants' responses would inevitably converge to a .5 probability of choosing either (they would be choosing between two identical stimuli) hence, in these circumstances, z must equal zero. However, a non-zero intercept would then make $z \neq 0$ so, for a model like this, it would be inappropriate to include one (Eisenhauer, 2003).

Because the stimuli come in complementary pairs (e.g., there is one stimulus that has a matched spectrum of 10-TET and an unmatched spectrum of 7-TET; there is a complementary stimulus with a matched spectrum of 7-TET and an unmatched spectrum of 10-TET), the mean value of the harmonicity predictor will be close to zero (any deviation from zero will be due to the random selection of stimuli). The spectral pitch similarity predictor does not have this property because it is also a function of the melodic intervals used; generally, it will have a value greater than 0 because matched spectra typically have greater expected pitch similarity than unmatched. This means that, of the two predictors, only spectral pitch similarity has the capacity to account for participants choosing a matched spectrum on more than 50% of occasions over all the stimuli.

Both predictors are, for each stimulus, the difference between two cosine similarity values, hence they both share the same dimensionless units and overall possible range of values, which is -1 to 1 (as mentioned earlier, all cosine similarity values for non-negative vectors like spectral pitch vectors fall between 0 and 1). This means their relative importance (unique effect size) can be ascertained from the relative sizes of their optimized coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$.

Model fitting, evaluation, and analysis. The model's parameters were iteratively optimized in MATLAB using the `fmincon` routine to maximize the likelihood of the model given the data (under the presumption that the numbers of matched spectra chosen are binomially distributed). As a nonlinear optimization, the resulting parameter values may produce a local, not the global, likelihood maximum. However, numerous parameter start values were chosen and the model typically optimized to the same values. The two optimized predictors (spectral pitch similarity and harmonicity) have low correlation ($r(108) = -.09$, $p = .334$) so there are no concerns with multicollinearity. A plot of the deviation residuals against predicted values confirmed the errors were randomly distributed with no apparent pattern.

Although the full model, as specified in (1), superficially appears to be a standard (generalized linear) logistic regression model, it is important to note that it is actually fully nonlinear. This is because the predictors ($f_{\Delta S}$ and $f_{\Delta H}$) are nonlinear with respect to the parameters ρ and σ, and these parameters are optimized simultaneously with the logistic weights β_1 and β_2. This means there is no simple way to calculate the degrees of freedom in the full model, so the standard χ^2 significance tests used for logistic regression models are not appropriate (and neither are standard information criteria such as AIC and BIC, which do not take account of the possible additional flexibility inherent in nonlinear parameterizations; Pitt, Myung, and Zhang 2002). To test model flexibility and generalizability, we use cross-validation, as detailed below.

The optimized parameter values, their standard errors, and statistical tests for the model are summarized in Table 2. The standard errors and confidence intervals were calculated from a Hessian matrix numerically estimated at the optimized parameter values. The 99.9% confidence intervals for the logistic weights are $\beta_1 \in [0.15, 9.40]$ and $\beta_2 \in [0.76, 8.70]$.

For a logistic model, there is no single equivalent to the R^2 used in linear regression to estimate a model's fit to the data (Zheng & Agresti, 2000). In Table 2, we report two straightforward fit statistics (both of which correspond to R^2 when applied to linear models): the *model-data squared correlation* is between the predicted numbers of matched spectra chosen and the numbers actually chosen (over the 110 stimuli, as depicted in Fig. 7); the *deviance* R^2 (also called the *Kullback-Leibler* R^2) gives the proportion of the maximal possible log-likelihood increase beyond the null model; that is, $R^2 = \frac{\log(L_{\text{mod}}) - \log(L_{\text{null}})}{\log(L_{\text{sat}}) - \log(L_{\text{null}})}$, where L_{sat}, L_{mod}, and L_{null} denote the likelihoods of the saturated, fitted, and null models, respectively (Cameron & Windmeijer, 1996, 1997).[7] The adjacent values subscripted with CV were obtained over ten runs of 10-fold cross-validation.

Figure 7 is a scatter plot, for all 110 stimuli, of the observed against predicted numbers of matched spectra chosen. Figure 8(b) shows the full model's predictions for all 110 stimuli. This can be usefully compared with the observed data shown in Figure 8(a) (which is the same as Figure 4 but without the stars). The individual contributions of the spectral pitch similarity and harmonicity predictors are shown in Figures 8(c) and 8(d).

The model-data squared correlation, the deviance R^2 value, and the scatter plot, indicate the model explains a sizeable proportion of the variance, or deviance, in the data. Under 10-fold cross-validation, these values drop (as would be expected), but only by small amounts. This indicates the model is not excessively flexible, and generalizes well beyond the data to which it is fitted.

[7] The *saturated* model is one with the best possible fit to the data—it has a parameter (predicted probability) for every stimulus (covariate pattern). The *null* model is usually specified as having only an intercept which, for our data, would predict a probability of .61 for every stimulus; but, because our model has no intercept, the appropriate null model is one that predicts a probability of .5 for all stimuli. Using the intercept-only null gives $R^2 = .55$, which is lower than when using the no-intercept null. This is because the former ignores the fitted model's ability to explain the upwards shift in the mean probability from the null hypothesis' .5 to the observed .61.

Table 2
Statistical analysis and evaluations of the full model and its parameters (the logistic part of the model does not include a constant term).

Optimized parameter	Value	SE	95% CI	
$\hat{\sigma}$ (smoothing)	10.53	1.09	8.40	12.66
$\hat{\rho}$ (roll-off)	0.58	0.18	0.23	0.94
$\hat{\beta}_1$ (spectral similarity weight)	4.78	1.41	2.02	7.53
$\hat{\beta}_2$ (harmonicity weight)	4.73	1.21	2.37	7.09
Overall model				
Model-data squared correlation	$r^2 = .64$		$r^2_{\text{CV}} = .61$	
Deviance R-squared	$R^2 = .68$		$R^2_{\text{CV}} = .65$	

Harmonicity and spectral pitch similarity are not correlated, so there are no concerns with multicollinearity in the full model, which means the estimates for the $\hat{\beta}_1$ and $\hat{\beta}_2$ coefficients should be reliable. Their 99.9% confidence intervals do not straddle zero, so we can be highly confident both spectral pitch similarity and harmonicity are having a positive influence on affinity. These two coefficients indicate the unique effect sizes of the spectral pitch similarity predictor and the harmonicity predictor (to remind, the former is the average spectral pitch similarity of successive intervals in the melody for the matched spectrum minus that of the unmatched spectrum; the latter is the harmonicity of the matched spectrum minus the harmonicity of the unmatched spectrum). More concretely, the $\hat{\beta}_1$ value indicates that, when the harmonicity predictor is held constant, a 0.1 increase in the spectral pitch similarity predictor results in a 61% increase in the odds of choosing a matched spectrum (because $e^{0.1\times4.78} = 1.61$). Similarly, the $\hat{\beta}_2$ value indicates that, when the spectral pitch similarity predictor is held constant, a 0.1 increase in the harmonicity predictor results in an 60% increase in the odds of choosing a matched spectrum ($e^{0.1\times4.73} = 1.60$). The similar values of these coefficients, and the previously mentioned lack of correlation between harmonicity and spectral pitch similarity, indicate that they have similar effect size and they are independent and complementary.

The optimized parameter values for the spectral roll-off ρ and smoothing width σ (0.58 and 10.53 cents, respectively) are reassuringly plausible in that they correspond with our prior expectations for their values. As discussed earlier, we would expect ρ to correspond approximately to the loudnesses of the partials, and for σ to have a value somewhere within or close to the just-noticeable frequency difference.

The partials in our stimuli had amplitudes of approximately $1/i$, where i is the partial number. According to Steven's power law, perceived loudness corresponds, approximately, to amplitude (pressure) to the power of 0.6, hence the loudness of each partial is approximately $1/i^{0.6}$, which is equivalent to $\rho = 0.6$ and is close to our

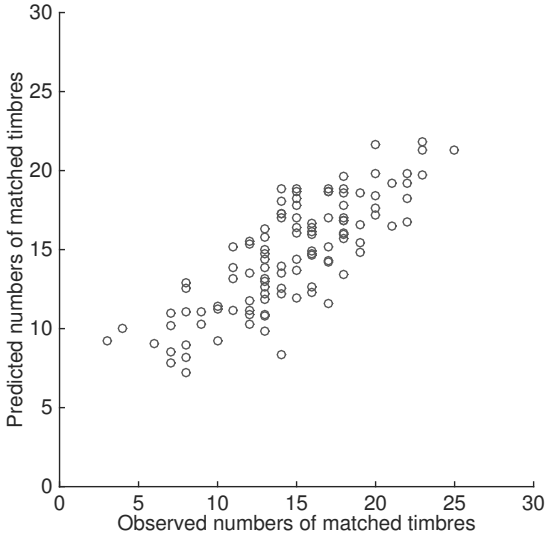


Figure 7. For all 110 observations, this scatter plot compares the observed numbers of matched spectra chosen by participants with those predicted by the full model. The 95% confidence interval for each data point has a mean of size of 9.2.

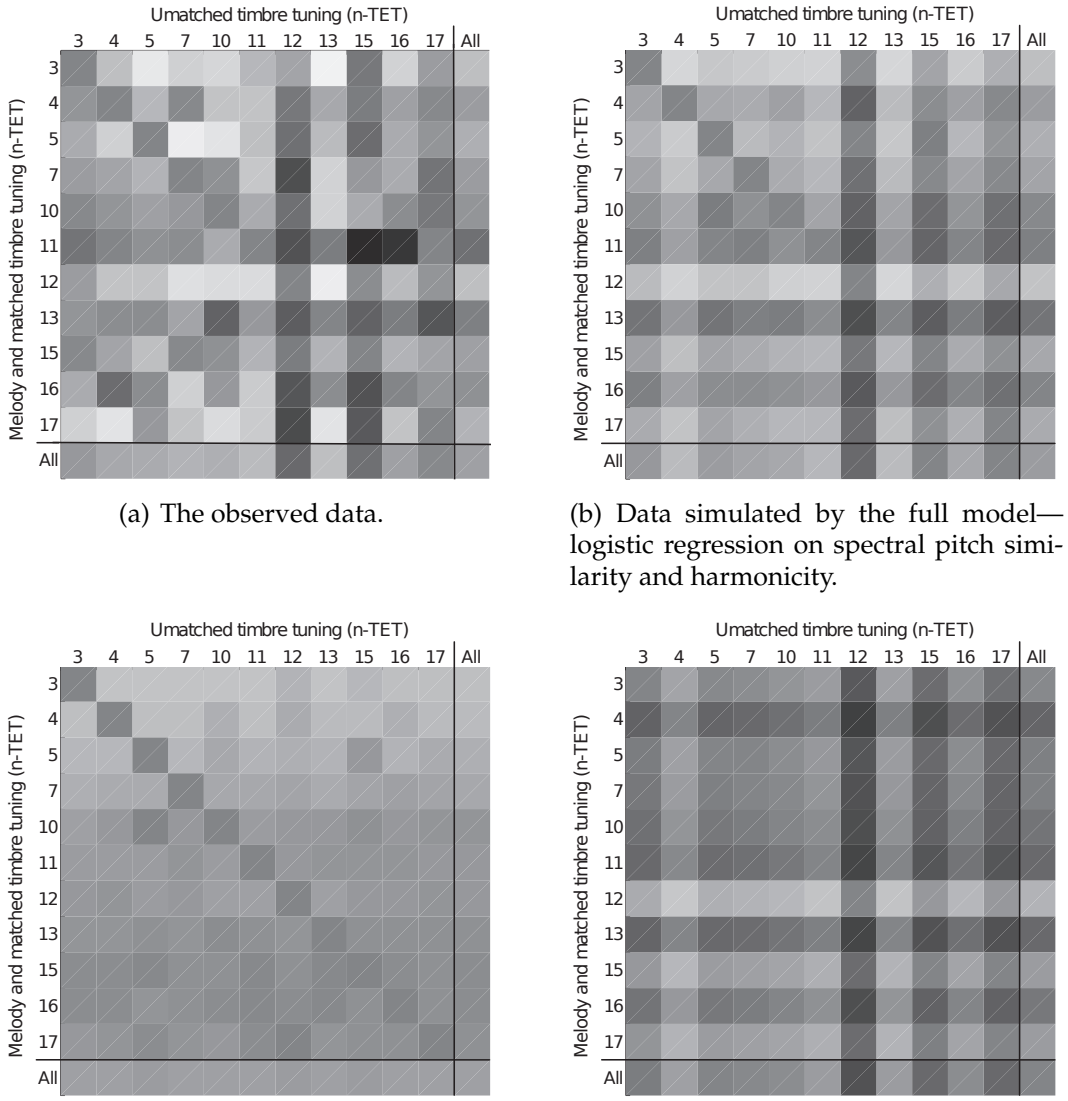
optimized value of 0.58.[8] For typical musical tones, which are harmonic complex tones and have stronger lower harmonics, this highlights the importance of intervals like the perfect fifth and perfect fourth whose low-numbered harmonics coincide (see Fig. 9).

Under experimental conditions, the frequency difference limen (just noticeable difference) is 3–13 cents between 125 and 6000 Hz (Moore, 1973), which would be modelled by an equivalent smoothing width. In an experiment like this, in which the stimuli are more explicitly musical, we would expect the standard deviation to be no smaller than this (Milne et al., 2011, App. A), the optimized value of approximately 10.5 cents meets these expectations. This value explains how intervals that are "imperfectly" tuned, in that no two partials perfectly coincide in frequency but still come close, can still have high affinity. For instance, the 12-TET perfect fifth is two cents smaller than the 3/2 frequency ratio where the second and third harmonics would perfectly coincide (assuming harmonic complex tones), but it is still typically regarded as a high-affinity interval, as predicted by our model (see Fig. 9).

Discussion

The results show that, in the context of this experiment, participants' ratings of the overall affinity of successive tones in a melody are positively affected, equally and independently, by both spectral pitch similarity and harmonicity. The stimuli had a large variety of harmonicities, scale tunings, pitch orderings, interval sizes, contour, tempo, articulation, and so forth, whilst still conforming with aspects of melodic structure that are widely exhibited in real-world music (e.g., prevalence of small steps over large, prevalences of certain ranges of tempo and articulation). This suggests

[8]Steven's law is a simplistic model of loudness, and is usually applied only to single sounds, but we use it here just to provide a quick "reality check" on the optimized value of ρ.



(a) The observed data.

(b) Data simulated by the full model—logistic regression on spectral pitch similarity and harmonicity.

(c) Data simulated by a logistic regression on just spectral pitch similarity.

(d) Data simulated by a logistic regression on just harmonicity.

Figure 8. The observed data and the modelled data.

that spectral pitch similarity and harmonicity will play a similar role in real-world perceptions of real-world music.

The experimental design eliminated any impact of interval familiarity on participants responses. This indicates that spectral pitch similarity is modelling a psychoacoustic process that affects interval preferences prior to knowledge of their typical prevalences or musical uses. The experimental design does not allow us to say the same about harmonicity, where preferences for tones with greater harmonicity may be down to acculturation. In future work, it would be interesting to test participants more familiar with inharmonic spectra (e.g., gamelan musicians or bell-ringers)—in such a situations, we may find that harmonicity has a reduced effect size.

Due to their musical expertise and experience, we would expect composers and performers to be particularly sensitive to the degree of affinity evoked by successive sounds. This is important because it is through the process of composition that psychoacoustic principles such as these become embedded into musical structure; for example, by favouring high affinity over low affinity through the use of sounds with high harmonicity (harmonic complex tones) and intervals with high spectral pitch similarity (e.g., perfect fifths). Once psychoacoustically motivated structures are common within a musical corpus, this further exemplifies, confirms, and stabilizes the musical meaning of such psychoacoustical relationships. In this way, psychoacoustic models such as these can shed light upon why certain musical structures are privileged, or how they are utilized, in a way that statistical models of familiarity cannot (Milne et al., 2015). We discuss more concrete examples of how spectral pitch similarity may have affected historical scale structures in the following subsection.

Another implication of this research is with regard to the composition and realization of novel experimental music. Previously, the notion of matching spectra and scales has been theoretically motivated on the basis of minimizing the sensory dissonance caused by the perception of rapid beating between simultaneously playing tones (e.g., Sethares 2005; Sethares et al. 2009). The research described in this paper shows that spectral matching can also be used to enhance the affinity of non-simultaneous tones. Indeed, it was our practical experience with Dynamic Tonality synthesizers—noticing, for example, how much more in-tune 5-TET melodies sound when the spectral tuning is matched—that first motivated this experiment. Having said that, it is also clear that spectra with partials close in frequency to the familiar harmonic template were typically preferred by our participants. This means that, in matching partials to a low n n-TET, one is often trading increased consonance and affinity between tones for decreased consonance within tones (bearing in mind that the latter may be a learned response).

A related implication is that we may be able to create musically compelling sequences where tension and release are modulated by spectral changes instead of, or in addition to, the pitch changes that form the focus of traditional Western music. This is not a new theory (it is part of the discourse behind electroacoustic music, where musical events or gestures can be envisaged as residing in a pitch and timbre space; e.g., Wishart 1983), but the model presented here may provide a way to more clearly estimate, for compositional and analytical purposes, the perceptual effects of such music.

Spectral Pitch Similarity as a Causal Influence on Scale Structure

The majority of pitched Western instruments have spectra whose partials follow a harmonic series (e.g., bowed string, wind instruments, and vocal vowels), or closely approximate one (e.g., plucked and hammered string instruments). Such spectra are also common in non-Western music and would have been found in any ancient music using wind-blown flutes, plucked strings, or singing. Figure 9 shows the spectral pitch similarity of pairs of tones with harmonic spectra separated by an interval

whose size is shown on the horizontal axis. Each tick corresponds to one 12-TET semitone, and a total range of just over one octave is covered.

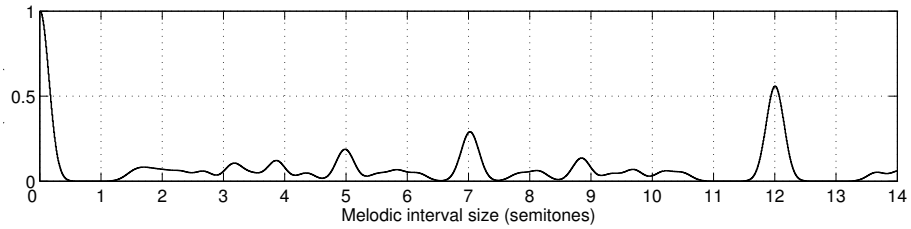


Figure 9. The spectral pitch similarity of pairs of harmonic complex tones with differing interval sizes (calculated with ρ and σ as optimized to the data). The graph bears a clear resemblance to the sensory dissonance charts of, for example, Plomp and Levelt (1965) and Sethares (2005), with maxima of modelled affinity at simple frequency ratios like 2/1, 3/2, 4/3, and so forth.

Clearly, the intervals with the highest spectral pitch similarity (other than the unison) are the octave and the perfect fifth and perfect fourth. There is significant empirical evidence that the octave is universally recognized as an interval with extremely high affinity (Deutsch 1977 and Woolhouse 2009 both cite numerous examples). As noted by Helmholtz (1877) (using different terminology), the high spectral pitch similarity of perfect fifths and perfect fourths tallies with historical evidence. For example, ancient Greek scales were typically based on conjunct and disjunct tetrachords. The two outer tones of a tetrachord span a perfect fourth (of frequency ratio 4/3) and, within this perfect fourth, lie two additional tones that could take on a wide variety of different tunings. The outer fourth was, however, always fixed. A second tetrachord was placed a whole-tone below the bottom note of the first tetrachord (i.e., a perfect fifth below the top note), so the entire octave was spanned to make a seven-tone scale. Typically, the two tetrachords had identical internal structure (Barbour, 1951), so the resulting scale was rich in high spectral pitch similarity perfect fourths and perfect fifths (it had at least four of each within the octave). This technique of scale construction might, therefore, be seen as a heuristic for creating high-affinity scales. Indeed, the bounding fourths potentially provide perceptually secure start and end points for a melody that traverses the more challenging tones in between. For an in-depth examination of the history, and mathematical, perceptual, and aesthetic properties of tetrachords, see Chalmers (1990) and Xenakis (1971), and for a discussion of the affinity (CDC-1) of the perfect fourth and fifth see Tenney (1988).

The diatonic and pentatonic scales, which are so ubiquitous to Western music are the richest in terms of perfect fifths and fourths (four of each in the former, six of each in the latter). This is because they are actually generated by a continuous chain of either of these intervals: there is no five-tone scale with more perfect fifths and fourths than the (anhemitonic) pentatonic, and no seven-tone scale with more perfect fifths and fourths than the diatonic. Such scales, therefore, maximize the number of the highest affinity (non-octave) intervals.

Conclusion

The model's fit to the experimental data, its ability to generalize over cross-validation, the plausibility of its parameter values, and the above observations of historical and contemporary musical structures, strongly support the two principal hypotheses that affinity is a monotonic function of spectral pitch similarity and harmonicity. The experimental design also supports the hypothesis that spectral pitch similarity is modelling a psychoacoustic process rather than one based on expectations driven by long-term memory.

There is no conceptual reason why the same (or similar) models could not also be applied to successions of chords, and to broader aspects of tonal functionality in both standard Western traditions and in experimental systems with unfamiliar timbres and tuning systems (Milne et al., 2015). For example, in other recent research we have demonstrated that spectral pitch similarity provides highly effective models of Krumhansl's (1982) tonal hierarchies (Milne et al., 2015), and of participants' ratings of the fit and similarity of major and minor triads (Milne & Holland, 2015). In both cases, the optimized smoothing width and roll-off parameters had values similar to those here.

We do not here seek to deny the important role of learning in determining musical expectancies and perceived fit (as evidenced by, e.g., Francès 1988; North and Hargreaves 1995; Schellenberg and Trehub 1999; Trehub et al. 1999; Pearce and Wiggins 2006), but these results suggest that psychoacoustic processes play a foundational role in determining the affinity of successive tones and, by extension, chords and other sounds.

Acknowledgements

Stefan Kreitmayer for assistance with the JavaScript parts of the Max patch, and Anthony Prechtl for building The Viking. This paper is based on work conducted at The Open University for a doctoral thesis (Milne, 2013). Aspects of the modelling and analysis used in this paper differ from the above thesis. We would also like to thank members of MARCS Institute for Brain, Behaviour and Development (notably, Roger Dean, Steffen Herff, and Kirk Olsen) for constructive criticism, as well as the anonymous reviewers who made exceptionally helpful comments and suggestions. Supplemental online material (audio examples, software, and raw data) is available from `http://msx.sagepub.com/content/by/supplemental-data`.

References

Balzano, G. J. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal*, *4*, 66–84.

Barbour, J. M. (1951). *Tuning and temperament: A historical survey*. East Lansing, Michigan: Michigan State College Press.

Bernstein, J. G., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *The Journal of the Acoustical Society of America*, *113*, 3323–3334.

Brown, J. C. (1992). Musical fundamental frequency tracking using a pattern recognition method. *The Journal of the Acoustical Society of America*, *92*, 1394–1402.

Cameron, A. C., & Windmeijer, F. A. G. (1996). *R*-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics*, *14*, 209–220.

Cameron, A. C., & Windmeijer, F. A. G. (1997). An *R*-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, *77*, 329–342.

Carey, N. (1998). *Distribution modulo one and musical scales* (Doctoral dissertation). University of Rochester.

Carey, N., & Clampitt, D. (1989). Aspects of well-formed scales. *Music Theory Spectrum*, *11*, 187–206.

Chalmers, J. (1990). *Divisions of the tetrachord* (L. Polansky & C. Scholz, Eds.). Frog Peak Music.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review*, *121*, 33–65.

Dahlhaus, C. (1990). *Studies on the origin of harmonic tonality* (R. O. Gjerdingen, Trans.). Oxford: Princeton University Press.

Deutsch, D. (1977). Memory and attention in music. In M. Critchley & R. A. Henson (Eds.), *Music and the brain: Studies in the neurology of music.* Southampton, UK: The Camelot Press.

DeWitt, L., & Crowder, R. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception and Psychophysics*, *41*, 73–84.

Eisenhauer, J. G. (2003). Regression through the origin. *Teaching Statistics*, *25*, 76–80.

Erlich, P. (2006). A middle path between just intonation and the equal temperaments, part 1. *Xenharmonikôn*, *18*, 159–199.

Helmholtz, H. L. F. (1877). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.). New York: Dover.

Huron, D. (2001, September). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*, 1–64.

Jerket, J. (2004, June). Music articulation in the organ. In *Proceedings of joint Baltic-Nordic acoustics meeting*. Mariehamn, Åland, Finland.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory parts 1 and 2. *The Journal*

of the Acoustical Society of America, *45*, 1451–1469.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346–374.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*, 334–368.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Berlin: Springer-Verlag.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*, 481–509.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*, 1035–1041.

Milne, A. J. (2013). *A computational model of the cognition of tonality* (Doctoral dissertation, The Open University, Milton Keynes, UK).

Milne, A. J., & Holland, S. (in press). Empirically testing voice-leading, Tonnetz, and spectral models of perceived harmonic distance. *Journal of Mathematics and Music*.

Milne, A. J., Laney, R., & Sharp, D. B. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception*, *32*, 364–393.

Milne, A. J., & Prechtl, A. (2008). New tonalities with the Thummer and The Viking. In A. Crossan & T. Kaaresoja (Eds.), *Proceedings of the 3rd International Haptic and Auditory Interaction Design Workshop* (Vol. 2, pp. 20–22). Jyväskylä, Finland.

Milne, A. J., Sethares, W. A., Laney, R., & Sharp, D. B. (2011). Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, *5*(1), 1–20.

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, *54*, 610–619.

Moore, B. C. J. (2005). *Introduction to the psychology of hearing*. London: Macmillan.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

Parncutt, R. (1992). *How "middle" is middle C? Terhardt's virtual pitch weight and the distribution of pitches in music.* (Unpublished)

Parncutt, R. (1994). Template-matching models of musical pitch and rhythm perception. *Journal of New Music Research*, *23*, 145–167.

Parncutt, R. (2011). The tonic as triad: Key profiles as pitch salience profiles of tonic triads. *Music Perception*, *28*, 333–365.

Parncutt, R., & Strasburger, H. (1994). Applying psychoacoustics in composition: "Harmonic" progressions of "nonharmonic" sonorities. *Perspectives of New Music*, *32*, 88–129.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472–491.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, *38*, 548–560.

Rahn, J. (2014). Pairs of interval classes in southeast Asian tunings. In *Third international conference on analytical approaches to world music*. London.

Roederer, J. G. (2008). *The physics and psychophysics of music: An Introduction* (4th ed.). Berlin: Springer.

Schneider, A. (1997). "Verschmelzung", tonal fusion, and consonance: Carl Stumpf revisited. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (Vol. 1317, pp. 117–13). Berlin: Springer.

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale* (2nd ed.). London: Springer Verlag.

Sethares, W. A., Milne, A. J., Tiedje, S., Prechtl, A., & Plamondon, J. (2009). Spectral tools for Dynamic Tonality and audio morphing. *Computer Music Journal*, *33*, 71–84.

Stevens, S. S. (1957). On the psychophysical law. *The Psychological Review*, *64*, 153–181.

Stumpf, C. (1890). *Tonpsychologie* (Vol. 2). Leipzig: Hirzel.

Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, *21*(153), 65–66.

Tenney, J. (1988). *A history of 'consonance' and 'dissonance'*. New York: Excelsior Music.

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception*, *1*, 276–295.

Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, *71*, 679–688.

Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, *6*, 383–396.

Wilson, E. (1975). *Letter to Chalmers pertaining to moments-of-symmetry/Tanabe cycle.* [PDF document]. (Retrieved from `http://www.anaphoria.com/mos.pdf`)

Wishart, T. (1983). *On sonic art*. London: Gordon and Breach.

Woolhouse, M. (2009). Modelling tonal attraction between adjacent musical elements. *Journal of New Music Research*, *38*, 357–379.

Xenakis, I. (1971). *Formalized music: Thought and mathematics in composition*. Bloomington and London: Indiana University Press.

Zheng, B., & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, *19*, 1771-1781.

Appendix A
Melody Generation: Model and Parameters

The model for generating the melodies was designed to emulate common melodic features in melodies. Some of these are defined for standard Western scales, and so need to be generalized to microtonal scales. The model comprises four parameterized probability distributions: two for actual pitch choices; two for pitch transitions. These were combined into a single first-order stochastic (Markov) matrix, from which the melodies were generated.

The first distribution is used to favour smaller over larger intervals. This was modelled by a triangular distribution over the size of the directed pitch interval between successive tones (measured in cents). The distribution was centred at zero and had a full-width at half-maximum of 1000 cents. This means the most probable directed pitch interval was zero (i.e., pitch repetition), while upwards and downwards pitch intervals become progressively less probable as they increase in size. They reach zero probability for intervals of size 1000 cents or greater (10 semitones, which is a minor seventh in the standard 12-TET diatonic scale).

The second distribution is used to provide an average pitch close to D♯4. This was modelled by a triangular probability distribution over the pitch height of each tone. The distribution was centred at D4 with a full width at half maximum of one octave. This means all pitches were higher than D3 and lower than D5.

The third distribution is used to generalize the notion of diatonic and chromatic to microtonal scales, and to favour diatonic pitches over chromatic. The pentatonic and diatonic scales are members of a broad class of scales called *well-formed* (Carey & Clampitt, 1989). Such scales contain no more than two step sizes, which are as distantly spaced as possible. Well-formedness has been suggested as a common organizational principle behind Western and non-Western scales (Carey & Clampitt, 1989; Rahn, 2014), and as a strategy for building musically intelligible microtonal scales (Carey, 1998; Erlich, 2006; Wilson, 1975). Each *n*-TET contains a unique set of well-formed scales; for example, in 11-TET, we can form a 7-note well-formed scale with steps of 2-2-1-2-1-2-1 (compare with the diatonic scale in 12-TET, which is 2-2-1-2-2-2-1).

All pitch classes in any *n*-TET can be generated by a pitch class interval called a *generator*. For example, all pitch classes in 12-TET can be produced by either a chain of semitones (e.g., C-C♯-D-D♯-E-F-F♯-G-G♯-A-A♯-B) or by a chain of perfect fifths (e.g., F-C-G-D-A-E-B-F♯-C♯-G♯-D♯-A♯) (Balzano, 1980). Each different *n*-TET implies a different set of possible generators (e.g., 7-TET can be generated by a 685.7 cents generator, 10-TET can be generated by 360 cents). Regardless of the *n*-TET, the notes in a well-formed scale are always a subsection of a generator chain; for example, the pentatonic scale is the chain of fifths (F-C-G-D-A), the diatonic scale is the chain of fifths (F-C-G-D-A-E-B)—both of which are subsections of the larger 12-note chain. This implies that within-well-formed-scale ("diatonic") note transitions typically use smaller directed generator distances (measured along the chain) than out-of-well-formed-scale ("chromatic") note transitions. Hence, to favour the former, a triangular distribution was used over the directed generator interval between

successive tones. The distribution was centred at zero and had a full-width at half-maximum of 6.5 generators. For example, in 12-TET, an ascending perfect fifth (or descending perfect fourth) implies a directed generator interval of 1; a descending perfect fifth (or ascending perfect fourth) implies a directed generator interval of -1; an ascending major second (or descending minor seventh) implies a directed generator interval of 2; a descending major second (or ascending minor seventh) implies a directed generator interval of -2, and so forth. The generators used for each of the 11 different n-TETs are shown in Table A1 (each n-TET typically has more than one distinct interval that can function as a generator, so the choice is arbitrary).

Table A1
Generator sizes (cents) for the different tunings.

3-TET	4-TET	5-TET	7-TET	10-TET	11-TET	12-TET	13-TET	15-TET	16-TET	17-TET
400	300	720	685.71	360	327.27	700	369.23	320	375	705.88

The fourth distribution is used to reduce the number of modulations between the scales produced in the previous step. This was modelled by a triangular distribution over the directed generator interval from the pitch class D. The distribution was centred at zero and had a full-width at half-maximum of 6 generators. For example, in 12-TET, the pitch class D has a directed generator interval (from D) of 0, A has a directed generator interval (from D) of 1, G has a directed generator interval of -1, E has a directed generator interval of 2, C has a directed generator interval of -2, and so forth. As before, directed generator intervals apply in an analogous fashion to the differing generators of differing n-TETs.

These probability distributions were used to populate first-order stochastic matrices giving the probabilities of each possible melodic transition. These matrices were entrywise multiplied and each row of the resulting matrix normalized to ensure probabilities in each row summed to unity. This final stochastic matrix was used to randomly generate the melodies. The Max patch embodying and utilizing this distribution can be downloaded from the supplemental online section.

The parameter values, given above, for the spreads of the four triangular probability mass functions, were initially chosen by reference to conventional musical practice and then refined to produce naturalistic sounding melodies. The precise values used are not critical so long as they fulfil their purpose, which is to avoid gross violations of typical melodic structure.

Appendix B
Tables of experimental data

Table B1

Ratio of number of matched spectra chosen and number of trials for each stimulus. Each row is a unique underlying and matched spectral tuning, each column is a unique unmatched spectral tuning.

	3-TET	4-TET	5-TET	7-TET	10-TET	11-TET	12-TET	13-TET	15-TET	16-TET	17-TET	All
3-TET		17/23	17/19	21/26	20/24	20/28	16/25	25/27	13/28	18/22	15/25	182/247
4-TET	15/26		19/27	12/23	19/25	18/24	13/28	17/26	10/21	13/21	13/25	149/246
5-TET	12/18	20/25		22/24	21/24	14/19	9/21	16/22	10/24	16/24	13/23	153/224
7-TET	15/25	14/22	16/23		14/25	23/30	8/28	18/22	10/17	16/24	9/20	143/236
10-TET	11/21	13/23	16/26	13/22		16/24	13/30	13/16	14/21	13/24	13/28	135/235
11-TET	11/25	12/24	11/20	12/22	14/21		8/26	13/27	3/23	4/20	12/24	100/232
12-TET	14/23	18/24	18/24	20/23	18/21	23/27		22/24	14/26	18/25	15/26	180/243
13-TET	15/26	13/24	12/22	19/30	8/22	16/27	8/23		8/22	12/25	7/22	118/243
15-TET	14/27	15/24	20/27	11/21	15/27	18/26	14/29	16/23		18/26	15/24	156/254
16-TET	14/21	7/17	15/28	17/21	17/29	18/23	7/22	13/24	7/23		15/26	130/234
17-TET	17/21	23/26	17/29	22/29	17/20	19/24	6/22	22/25	8/24	18/24		168/244
All	138/233	152/232	161/245	169/241	163/238	185/252	102/254	175/236	97/229	146/235	127/243	1615/2638

Table B2

As above, but expressed in decimal format.

	3-TET	4-TET	5-TET	7-TET	10-TET	11-TET	12-TET	13-TET	15-TET	16-TET	17-TET	All
3-TET		0.739	0.895	0.808	0.833	0.714	0.640	0.926	0.464	0.818	0.600	0.737
4-TET	0.577		0.704	0.522	0.760	0.750	0.464	0.654	0.476	0.619	0.520	0.606
5-TET	0.667	0.800		0.917	0.875	0.737	0.429	0.727	0.417	0.667	0.565	0.683
7-TET	0.600	0.636	0.696		0.560	0.767	0.286	0.818	0.588	0.667	0.450	0.606
10-TET	0.524	0.565	0.615	0.591		0.667	0.433	0.813	0.667	0.542	0.464	0.574
11-TET	0.440	0.500	0.550	0.545	0.667		0.308	0.481	0.130	0.200	0.500	0.431
12-TET	0.609	0.750	0.750	0.870	0.857	0.852		0.917	0.538	0.720	0.577	0.741
13-TET	0.577	0.542	0.545	0.633	0.364	0.593	0.348		0.364	0.480	0.318	0.486
15-TET	0.519	0.625	0.741	0.524	0.556	0.692	0.483	0.696		0.692	0.625	0.614
16-TET	0.667	0.412	0.536	0.810	0.586	0.783	0.318	0.542	0.304		0.577	0.556
17-TET	0.810	0.885	0.586	0.759	0.850	0.792	0.273	0.880	0.333	0.750		0.693
All	0.592	0.655	0.657	0.701	0.685	0.734	0.402	0.742	0.424	0.621	0.523	0.612

Table B3

Two-tailed p-values (exact binomial) giving the probability of the above data occurring under the null hypothesis that matched spectra are chosen with a probability of .5.

	3-TET	4-TET	5-TET	7-TET	10-TET	11-TET	12-TET	13-TET	15-TET	16-TET	17-TET	All
3-TET		0.035	0.001	0.002	0.002	0.036	0.230	0.000	0.851	0.004	0.424	0.000
4-TET	0.557		0.052	1.000	0.015	0.023	0.851	0.169	1.000	0.383	1.000	0.001
5-TET	0.238	0.004		0.000	0.000	0.064	0.664	0.052	0.541	0.152	0.678	0.000
7-TET	0.424	0.286	0.093		0.690	0.005	0.036	0.004	0.629	0.152	0.824	0.001
10-TET	1.000	0.678	0.327	0.523		0.152	0.585	0.021	0.189	0.839	0.851	0.026
11-TET	0.690	1.000	0.824	0.832	0.189		0.076	1.000	0.000	0.012	1.000	0.042
12-TET	0.405	0.023	0.023	0.000	0.001	0.000		0.000	0.845	0.043	0.557	0.000
13-TET	0.557	0.839	0.832	0.200	0.286	0.442	0.210		0.286	1.000	0.134	0.700
15-TET	1.000	0.307	0.019	1.000	0.701	0.076	1.000	0.093		0.076	0.307	0.000
16-TET	0.189	0.629	0.851	0.007	0.458	0.011	0.134	0.839	0.093		0.557	0.102
17-TET	0.007	0.000	0.458	0.008	0.003	0.007	0.052	0.000	0.152	0.023		0.000
All	0.006	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.024	0.000	0.521	0.000